Ⅲ. 표본설계 개요

1. 모집단의 정의 및 표본추출틀

(1) 목표모집단 : 사업체* 정의에 부합되는 대한민국내 '전문, 과학 및 기술 서비스업'을 영위하는 사업체

* 사업체의 정의 : 일정한 물리적 장소 또는 일정한 지역 내에서 하나의 단일 또는 주된 경제활동에 독립적으로 종사하는 기업체 또는 기업체를 구성하는 부분단위

(2) 조사모집단 : 제9차 한국표준산업분류상 'M 전문, 과학 및 기술 서비스업' 중 일부 업종(70. 연구개발업, 7151 회사본부)을 제외한 모든 사업체

(3) 표본추출틀 : 조사모집단을 구성하는 '12년 기준 전국사업체조사 사업체 명부(76,622개)

2. 표본설계

(1) 부차모집단 : 시도 및 산업세세분류

(2) 표본추출방법 : 층화계통추출방법

(3) 층화방법 : 부차모집단별로 다음과 같이 층화

○ 1차 층화 : 전수층, 표본층

- 전수층

・전수업종

‥기타 회계관련 서비스업(71209), 비금융 지주회사(71520)

‥전국ㆍ산업세세분류별 사업체수가 30개 이하인 업종

‥시도ㆍ산업세세분류별 20개 이하인 업종

・전수경계점 이상

‥매출액 100억 이상인 사업체

‥종사자수 50인 이상인 사업체

‥$(y_{i+1}-y_i) \geqq (3\times\sigma_y)$인 경우, 매출액이 상대적으로 큰 사업체

- 표본층 : 전수층 사업체를 제외한 나머지 사업체

○ 2차 층화 : 종사자규모층(1~4인, 5~9인, 10~19인, 20~49인, 50인 이상)

(4) 표본규모결정

○ 배분방법 : 네이만 배분법(neyman allocation)

○ 부문별 목표 상대표준오차(CV, Cofficient of Variation)
- 산업중분류별 목표 CV 1.1% 이내
- 시도별 목표 CV 1.5% 이내
- 산업세세분류별 목표 CV 3.2% 이내
- 시도 및 산업세분류별 목표 CV 5.1% 이내
- 시도 및 산업세세분류별 목표 CV 6.6% 이내

○ 표본수 계산 공식

· $n_{hi} = \dfrac{\sum_{j=1}^{5} N_{hij}^2 \cdot S_{r,\,hij}^2 / w_{hij}}{cv^2 \cdot Y_{hi}^2 + \sum_{j=1}^{5} (N_{hij} \cdot S_{r,\,hij}^2)}$

·· $w_{hij} = \dfrac{S_{r,\,hij} \cdot N_{hij}}{\sum_{j=1}^{5} S_{r,\,hij} \cdot N_{hij}}$

·· $S_{r,\,hij} = \dfrac{\sum_{l}^{N_{hij}} (y_{hij} - R_{hij} x_{hij})^2}{N_{hij} - 1}$, 여기서 $R_{hij} = \dfrac{\sum_{l}^{N_{hij}} y_{hijl}}{\sum_{l}^{N_{hij}} x_{hijl}}$

· $n_{hij} = n_{hi} \cdot w_{hij}$

첨자 h : 시도
i : 산업세세분류
j : 1,2…5, 종사자규모층

변수 n : 표본크기
N : 모집단크기
Y : 주특성치(매출액)
x : 보조변수(종사자수)
cv : 목표 상대표준오차
w : 층별 표본수 배분 비율

(5) 표본사업체 선정 : 부차모집단 및 종사자규모 층별로 사업체를 매출액 내림차순으로 정렬한 후 계통추출(중앙값 계통추출)

3. 총합 추정

○ 사후층화 비추정방법 : 표본설계시 전년기준 표본추출틀을 사용함에 따라 추정단계에서 최신 자료로 사후층 형성

- 주요 특성변수인 매출액과 상관관계가 높은 종사자수의 비(ratio)를 이용하여 모집단 총합을 추정

○ 분리 비(separate ratio) r 추정 : $\hat{r}_{hij} = \dfrac{\sum_{l}^{{}_{s}n_{hij}} y_{hijl}}{\sum_{l}^{{}_{s}n_{hij}} x_{hijl}}$

- 종사자층별 총합 추정 : $\hat{\tau}_{y,hij} = \sum_{l}^{n_{hij}} y_{hijl} + \hat{r}_{hij} \times {}_{s}T_{x,hij}$

- 시도별 총합 추정 : $\hat{\tau}_{y,h} = \sum_{i}\sum_{j} \hat{\tau}_{y,hij}$

- 전국 총합 추정 : $\hat{\tau}_{y} = \sum_{h} \hat{\tau}_{y,h}$

첨자 c : 전수층 사업체
s : 표본층 사업체
h : 시도
i : 산업세세분류
j : 종사자규모층
l : 개별사업체

변수 x : 종사자수
y : 매출액

4. 표본분산 추정

○ 분리 비 추정 분산식

- 종사자층별 총합 $\hat{\tau}_{y,hij}$의 분산 :

$$\hat{V}(\hat{\tau}_{y,hij}) = {\tau_{x,hij}}^2 \hat{V}(r_{hij}) = \frac{{}_sN_{hij}({}_sN_{hij} - {}_sn_{hij})}{{}_sn_{hij}} {\hat{s}_{r,hij}}^2$$

$$\text{여기서, } \hat{s}^2_{r,hij} = \frac{\sum_{l}^{{}_sn_{hij}} (y_{hijl} - \hat{r}_{hij} x_{hijl})^2}{{}_sn_{hij} - 1}$$

- 시도별 총합 $\hat{\tau}_{y,h}$의 분산 : $\hat{V}(\hat{\tau}_{y,h}) = \sum_i \sum_j \hat{V}(\hat{\tau}_{y,hij})$

- 전국 총합 $\hat{\tau}_y$의 분산 : $\hat{V}(\hat{\tau}_y) = \sum_h \hat{V}(\hat{\tau}_{y,h})$

○ 표준오차 : $SE(\hat{\tau}_y) = \sqrt{\hat{V}(\hat{\tau}_y)}$

○ 상대표준오차 : $CV(\hat{\tau}_y) = \frac{SE(\hat{\tau}_y)}{\hat{\tau}_y} \times 100$

○ 100(1-α)% 신뢰구간 : $\hat{\tau}_y \pm z_{\alpha/2} SE(\hat{\tau}_y)$

Ⅲ. Outline of sample design

1. Definition of population and sample frame

(1) **Target population** : Domestic establishments* conducting professional, scientific and technical activities(section M)

* The establishment is defined as an enterprise or part of an enterprise that is situated in a single location and in which only a single(non-ancillary) productive activity is carried out or in which the principal productive activity accouts for most of the value added.

(2) **Survey population** : All establishments classified into section M. professional, scientific and technical activities of KSIC Rev.9 except for establishments corresponding to 70 division(research and development) and 7151 class(activities of head offices)

(3) **Sample frame** : 76,622 establishments consisting of the survey population from the 2012 census on establishments

2. Sample design

(1) Sub-population : Province and sub-classes of industrial classification

(2) Sampling method : Stratified systematic sampling method

(3) Stratified method : Stratified by sub-population

○ Primary stratification : take-all stratum, take-some stratum

- Take-all stratum

· Complete survey section

·· Other accounting book-keeping and auditing activities(71209), Non-financial holding companies(71520)

·· **Survey section whose the** establishments number is 30 and less in a population by sub-classes of industry & whole country

·· **Survey section whose the** establishments number is 20 and less in a population by sub-classes of industry & province were completely surveyed

· Take-all threshold more

·· Establishments with 10 billion won and more in annual sales

·· Establishments with 50 and more in the number of workers

·· Application of $(y_{i+1} - y_i) \geqq (3 \times \sigma_y)$, establishments with relatively high annual sales

- Take-some stratum : Establishments excluding take-all stratum

○ Secondary stratification : size of workers stratum(1~4persons, 5~9, 10~19, 20~49, 50 persons and more)

(4) Determining the sample size

○ Allocation method : neyman allocation

○ Target Coefficient Variation(CV) by sector

- Within 1.1% of CV for divisions of industry
- Within 1.5% of CV for cities and provinces
- Within 3.2% of CV for sub-classes of industry
- Within 5.1% of CV for cities and classes of industry
- Within 6.6% of CV for cities and sub-classes of industry

○ Calculating the Sample size

· $$n_{hi} = \frac{\sum_{j=1}^{5} N_{hij}^2 \cdot S_{r,\,hij}^2 / w_{hij}}{cv^2 \cdot Y_{hi}^2 + \sum_{j=1}^{5} (N_{hij} \cdot S_{r,\,hij}^2)}$$

·· $$w_{hij} = \frac{S_{r,\,hij} \cdot N_{hij}}{\sum_{j=1}^{5} S_{r,\,hij} \cdot N_{hij}}$$

$$\cdot\cdot\ S_{r,hij} = \frac{\sum_{l}^{N_{hij}} (y_{hij} - R_{hij} x_{hij})^2}{N_{hij} - 1}, \quad \text{where } R_{hij} = \frac{\sum_{l}^{N_{hij}} y_{hijl}}{\sum_{l}^{N_{hij}} x_{hijl}}$$

$$\cdot\ n_{hij} = n_{hi} \cdot w_{hij}$$

subscript h : district
i : sub-classes of industry
j : 1,2…5, stratum by size of workers

variable n : Total sample size
N : Population
Y : pimary characteristics(annual sales)
x : Auxiliary characteristics(number of workers)
cv : Target relative standard error
w : Allocation rate of sample size for stratum

(5) Selecting the sample establishment

Annual sales of establishments were descending sorts by subpopulation and size of workers stratum. Then establishments were selected through systematic sampling

3. Estimate of total amount

○ Post stratified ratio estimation

- Estimaing total population amount using ratio of number of workers that highly correlating with annual sales, the primary characteristics.

○ Estimates of separate ratio : $\hat{r}_{hij} = \frac{\sum_{l}^{s^{n_{hij}}} y_{hijl}}{\sum_{l}^{s^{n_{hij}}} x_{hijl}}$

- Estimates by size of workers stratum : $\hat{\tau}_{y,hij} = \sum_{l}^{{}_{c}n_{hij}} y_{hijl} + \hat{r}_{hij} \times {}_{s}\tau_{x,hij}$

- Estimates by district : $\hat{\tau}_{y,h} = \sum_{i}\sum_{j} \hat{\tau}_{y,hij}$

- Estimates of whole country : $\hat{\tau}_{y} = \sum_{h} \hat{\tau}_{y,h}$

subscript	c	: take-all stratum	variable	x	: number of workers
	s	: take-some stratum		y	: annual sales
	h	: district			
	i	: sub-classes of industry			
	j	: stratum by size of workers			
	l	: individual establishment			

4. Sampling variance

○ Variation estimation of separate ratio

- Variation of total amount by size of workers stratum :

$$\hat{V}(\hat{\tau}_{y,hij}) = \tau_{x,hij}^{\,2}\ \hat{V}(r_{hij}) = \frac{{}_{s}N_{hij}({}_{s}N_{hij} - {}_{s}n_{hij})}{{}_{s}n_{hij}}\ \hat{s}_{r,hij}^{\,2}$$

$$\text{where, } \hat{s}_{r,hij}^{\,2} = \frac{\sum_{l}^{{}_{s}n_{hij}} (y_{hijl} - \hat{r}_{hij} x_{hijl})^2}{{}_{s}n_{hij} - 1}$$

- Variation of total amount by district : $\hat{V}(\hat{\tau}_{y,h}) = \sum_{i}\sum_{j} \hat{V}(\hat{\tau}_{y,hij})$

- Variation of whole country : $\hat{V}(\hat{\tau}_{y}) = \sum_{h} \hat{V}(\hat{\tau}_{y,h})$

○ **Standard error** : $SE(\hat{\tau_y}) = \sqrt{\hat{V}(\hat{\tau_y})}$

○ Relative **standard error** : $CV(\hat{\tau_y}) = \frac{SE(\hat{\tau_y})}{\hat{\tau_y}} \times 100$

○ 100(1-α)% confidence interval : $\hat{\tau}_y \pm z_{\alpha/2} SE(\hat{\tau}_y)$